

Probabilistic speed-density relationship for pedestrian traffic: a data-driven approach

Marija Nikolić ^{*} Michel Bierlaire ^{*} Bilal Farooq [†]

April 11, 2015

Report TRANSP-OR 150411
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
`transp-or.epfl.ch`

^{*}Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Switzerland, {marija.nikolic, michel.bierlaire}@epfl.ch

[†]Département des génies civil, géologique et des mines, Polytechnique Montréal, Canada, {bilal.farooq@polymtl.ca}

Abstract

We propose a probabilistic modeling approach to represent the speed-density relationship of pedestrian traffic. The approach is data-driven, and it is motivated by the presence of high scatter in the raw data that we have analyzed. We show the validity of the proposed approach, and its superiority compared to deterministic approaches from the literature using a dataset collected from a real scene and another from a controlled experiment.

Keywords: speed-density relationship, probabilistic model, individual trajectories, Voronoi tessellations, statistical validation

1 Introduction

Understanding, reproducing and forecasting phenomena that characterize pedestrian traffic is necessary in order to provide services related to pedestrian safety and convenience. This becomes of utmost importance in areas of high congestion, which is a growing problem of many public spaces (transportation hubs, shopping malls, large sports and cultural events, etc.) Congestion in pedestrian-oriented facilities represents a phenomenon with a negative impact on pedestrian dynamics. It prevents pedestrians from achieving efficient movements and may lead to an increase in travel time, delays and potential collisions among pedestrians. Because of the complex and heterogeneous patterns in pedestrian flows, a simple application of a particular policy may lead to inefficient and costly trial-and-error solutions.

Data collection for pedestrian flow and behavior analysis used to be particularly cumbersome. Typically, manual counting methods (on-site or on videos) and surveys distributed to randomly selected individuals were the main sources of data. Nowadays, automatic pedestrian detection and tracking methods have evolved tremendously, allowing for more comprehensive analyses to be conducted as well (Bauer et al., 2009).

Using a direct analogy with vehicular traffic, the main stream of the literature characterizes pedestrian traffic with three fundamental quantities, that is density (k), speed (v) and flow (q), as well as deterministic relationships among them. Density (in ped/m²) is the number of pedestrians present in an area at a given moment in time; speed (in m/s) is the mean speed of pedestrians which may be averaged over space or over time; and

flow (in ped/ms) refers to the number of pedestrians passing a cross section of an area per unit of time (Daamen, 2004). The relationships between density and flow, density and speed, and flow and speed is referred to as the *fundamental diagram* (see Weidmann, 1993; Daamen, 2004).

In this paper, we exploit data collected from the train station in Lausanne, Switzerland, as well as data collected from a controlled experiment by the Technical University of Delft (Daamen and Hoogendoorn, 2003). The empirical analysis of these pieces of data rules out the use of a unique deterministic fundamental diagram, due to a high scatter in the data. This scatter may be explained by the behavioral heterogeneity of pedestrians, as documented in the literature. Individuals with different ages, health conditions, trip purposes, with or without luggage, or walking in a group or alone, may behave differently.

A possible approach to capture this complex phenomenon consists in modeling explicitly each type of behavior at the disaggregate level. It would allow to test various behavioral hypotheses, at the expense of the collection of a great deal of disaggregate behavioral data. We propose in this paper an alternative approach, based on an aggregate representation of the pedestrian traffic (consistently with the fundamental diagram approach mentioned above), and directly derived from the data. The observed scatter is captured by preferring probabilistic models instead of deterministic ones.

The structure of the paper is as follows. A review of related research from the literature is provided in Section 2. Section 3 describes the two case studies from Lausanne and Delft mentioned above. In Section 4, we formally define the variables involved in the model, that is density and speed indicators for pedestrian traffic. Section 5 presents the empirical analysis of the two case studies and emphasizes the limitations of the state of the art approaches on these concrete examples. In Section 6, we introduce the specification of the probabilistic speed-density model. Section 7 and Section 8 illustrate the model on the two case studies. Parameter estimation, model validation and comparisons with the existing models are discussed in details. Finally, Section 9 summarizes the outcomes of the proposed methodology and determines future research directions.

2 Literature review

This section is organized into two parts. The first focuses on models from the vehicular traffic theory, that are relevant for pedestrian as well. The second deals with models for the pedestrian traffic.

2.1 Vehicular traffic

The fundamental relation between spacing (the inverse of density) and speed was first introduced by Greenshields et al. (1935) in a form of a simple linear equation. Since then there have been many studies that were aimed at improving this relationship. A comprehensive review of the models proposed in this field is given in Wang et al. (2009). Some of the established deterministic empirical relationships are listed in Table 1, where v_f is the free-flow speed, v_0 is the average travel speed in stop-and-go conditions, k_j is the jam density, k_c is the critical density, and λ , θ , θ_1 and θ_2 are the parameters.

Source	Specification	Parameters
Greenshields et al. (1935)	$v(k) = v_f \left(1 - \frac{k}{k_j}\right)$	v_f, k_j
Underwood (1961)	$v(k) = v_f \exp\left(-\frac{k}{k_c}\right)$	v_f, k_c
Newell (1961)	$v(k) = v_f \left(1 - \exp\left(-\frac{\lambda}{v_f} \left(\frac{1}{k} - \frac{1}{k_j}\right)\right)\right)$	v_f, k_j, λ
Drake et al. (1967)	$v(k) = v_f \exp(-\theta k^2)$	v_f, θ
Wang et al. (2009)	$v(k) = v_0 + \frac{v_f - v_0}{(1 + \exp(\frac{k - k_c}{\theta_1}))^{\theta_2}}$	$v_0, v_f, k_c, \theta_1, \theta_2$
Units: k [veh/km], v [km/h]		

Table 1: Deterministic fundamental relationships - vehicular traffic

In several recent studies, it has been recognized that plots of speed-density data are usually widely scattered. Researchers addressed this scatter through the modeling of macroscopic flow characteristics dependent on drivers' characteristics (Jabari et al., 2014), through the implementation of multiple driver and vehicle classes (van Wageningen-Kessels, 2013) or through the probabilistic extension of the existing macroscopic relations (Wang et al., 2009). These models explain the data better at the macroscopic level (Jabari et al., 2014).

2.2 Pedestrian traffic

In the context of pedestrian traffic, both linear and nonlinear speed-density models have been proposed, as reported in Table 2, where v_f is the free flow speed, k_j the jam density, and θ and γ are parameters. The linearity of the speed-density relationship has long been questioned for both vehicular and pedestrian flows (Daamen, 2004). An alternative specification has been proposed by Tregenza (1976) where speed decreases exponentially with the increase in density, whereas Weidmann (1993) proposed the so-called Kladek-formula, with a double S-form. The exponential specifications of the relationship appeared to be better for describing the behavior of pedestrian walking speed (Cheah and Smith, 1994). In comparison to fundamental relationships from vehicular traffic, the relationship proposed by Weidmann (1993) corresponds to the model proposed by Newell (1961), while the relationship proposed by Tregenza (1976) can be regarded as the generalization of the model proposed by Underwood (1961). Rastogi et al. (2013) have shown that the speed-density relationship of pedestrian flow on sidewalks also follows the model presented in Underwood (1961).

The proposed relationships clearly differ in terms of functional form, but also in terms of the values of their parameters and supports. For instance, jam density (the maximum density achieved under congestion) goes from 3.8 ped/m² to 10 ped/m², the reported critical density (the maximum density achievable under free flow) ranges from 1.7 ped/m² to 7 ped/m² (Seyfried et al., 2010) and the mean of the free-flow speed estimated in different studies is 1.34 m/s while its standard deviation is 0.37 m/s (Daamen, 2004). The researchers have suggested several explanations for these deviations some of which can be attributed to the cultural differences, the differences between pedestrian facilities and the effects of the environment, flow composition, measurement methods, etc. (Seyfried et al., 2010).

The findings from several studies (Cheung and Lam, 1998; Daamen et al., 2005; Steffen and Seyfried, 2010), question the deterministic approach. They indeed report a significant scatter in the empirical speed-density relationship. The observed scatter is not possible to predict by the proposed deterministic models. Cheung and Lam (1998) have reported different distributions of the speed data observed for various ranges of density. In this study, speeds are less evenly distributed for lighter traffic conditions,

Source	Specification	Parameters
Older (1968) Navin and Wheeler (1969) Fruin (1971) Tanaboriboon et al. (1986) Lam et al. (1995)	$v(k) = v_f - \theta k$	v_f, θ
DiNenno (2002)	$v(k) = v_f - v_f \theta k$	v_f, θ
Tregenza (1976)	$v(k) = v_f \exp\left(-\left(\frac{k}{\theta}\right)^\gamma\right)$	v_f, γ, θ
Weidmann (1993)	$v(k) = v_f \left\{1 - \exp\left(-\gamma \left(\frac{1}{k} - \frac{1}{k_j}\right)\right)\right\}$	v_f, k_j, γ
Rastogi et al. (2013)	$v(k) = v_f \exp\left(-\frac{k}{\theta}\right)$	v_f, θ
Units: $k[\text{ped}/\text{m}^2], v[\text{m}/\text{s}]$		

Table 2: Deterministic fundamental relationships - pedestrian traffic

which is explained by higher freedom that pedestrians have in controlling their movements. This indicates that in addition to density, other factors are likely to influence the speed of pedestrians. Weidmann (1993) has empirically shown that the trip purpose of pedestrians represents one of the relevant factors. According to this study free-flow speed of shopping pedestrians is 1.04 m/s, it is 1.45 m/s for commuters and 0.99 m/s for tourists. The speed of pedestrians appears to be affected by the age and the gender as well. According to Bowman and Vecellio (1994), the walking speed of pedestrians who are 60 years old and older is significantly lower than for the rest of the adult population. Weidmann (1993) has reported that children (under 12 years) are not capable of attaining the same speed as adults. According to the same study, walking speed of men is found to be 1.41 m/s, whereas for women it is lower (1.27 m/s).

Existing models are not designed to capture these complex aspects. This is where our study makes a contribution. We propose a probabilistic speed-density relationship that is able to implicitly account for the heterogeneity of pedestrian flows.

3 Case studies

The motivation of this research comes from the analysis of two real datasets, that we use below to illustrate and validate our approach.

3.1 Lausanne train station

The first dataset is collected in a pedestrian underpass of the train station of Lausanne, Switzerland. Figure 1 shows the layout of the studied area. It covers approximately 685 m². The underpass is frequently used especially during the morning and afternoon peak hours since it connects the exterior of the train station to the main platforms. It also acts as a connection between mostly residential south and the center of the city in the north.

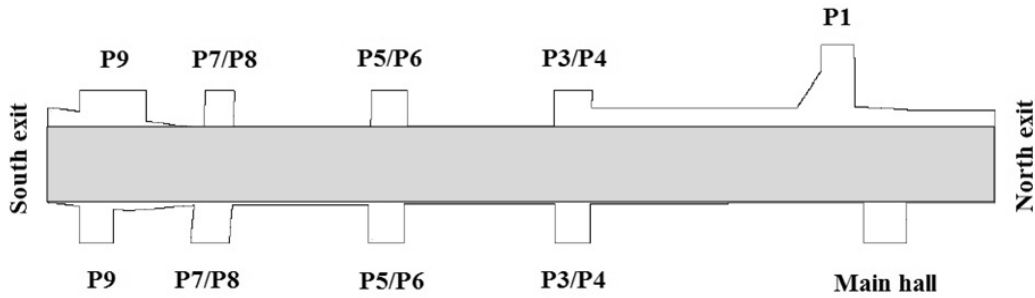


Figure 1: Lausanne train station - pedestrian underpass West

To collect the raw data, a large-scale network of smart sensors has been deployed in the station. The underlying technology is based on infrared and depth sensors that detect silhouettes and track each pedestrian in the scene covered by the network. The tracking engine uses a sparsity driven framework (Alahi et al., 2011; Alahi et al., 2014) to link detected pedestrians over the network of sensors.

It results in a dataset of 25,603 trajectories, collected in a time period between 07:00 and 08:00 on February 12, 13, 14, 15 and 18 of 2013. The temporal resolution of every trajectory is 10 to 25 points per second and it has been processed to obtain the position of every pedestrian in the scene every second. The average length of the trajectories is 78 meters and the duration of pedestrians' stay in the underpass ranges from 15 seconds to 2.2 minutes.

Note that we have selected only trajectories collected in the shaded area shown in Figure 1, referring to a corridor. The trajectories from the ramps and stairs (denoted as P1-P9) are not considered in this study. Indeed, as explained by Daamen (2004) and Weidmann (1993), the walking behavior and, therefore, the speed-density relationship, varies with the type of infrastructure.

In the rest of the paper, we refer to this case study as the Lausanne case study.

3.2 Controlled experiment

The second set of data has been collected during a controlled experiment at the Technical University of Delft in the Netherlands (Daamen and Hoogendoorn, 2003). The individuals participating in the experiment were instructed to walk along a corridor that is 10 meters long and 4 meters wide, at a normal speed, and to pass through a bottleneck of 1 meter in width (see Figure 2, where individuals walk from right to left).



Figure 2: Narrow bottleneck experiment (Daamen and Hoogendoorn, 2003)

The scene was filmed from top by digital cameras. The individual trajectories were extracted from the digital video sequences.

The experiment lasted about 15 minutes. A total of 1,123 trajectories were collected, where the position of each individual is available every 0.1s. The average length of the trajectories is similar inside and upstream of the bottleneck and it is approximately 5 meters. The average travel time of the trajectories upstream of the bottleneck is 10 seconds, whereas inside the bottleneck it is lower (approximately 5 seconds).

Note that we have selected trajectories collected in the rectangular area (5 meters long and 4 meters wide) upstream of the bottleneck (Figure 2).

As explained by Duives et al. (2014, Figure 4), this is where the variability is observed.

In the rest of the paper, we refer to this case study as the Delft case study.

4 Density and speed indicators

We present here the assumptions related to the quantities involved in our analysis. The *trajectory* of pedestrian i is a curve in space and time, that is

$$\mathbf{p}_i(t) = (x_i(t), y_i(t), t), \quad (1)$$

where time t spans the horizon of the analysis $[t_0, t_f]$, and $x_i(t)$ and $y_i(t)$ are the coordinates of the position of pedestrian i at time t in a given system of coordinates (typically, we express time in seconds, and use an orthonormal basis for the spatial dimensions).

In practice, the pedestrian trajectory data is collected through an appropriate tracking technology (e.g. Daamen and Hoogendoorn, 2003; Alahi et al., 2011). In this case, the time is discretized and the trajectory is described as a finite collection of triples

$$\mathbf{p}_{is} = (x_{is}, y_{is}, t_s), \quad (2)$$

where $t_s = (t_0, t_1, \dots, t_f)$ corresponds to the available sample. We assume that the position of each pedestrian is known at each time t_s of the discretization.

Different measurement methods have been proposed in the literature in order to obtain density and speed indicators from pedestrian trajectories. For a comprehensive analysis of several measurement methods and their influence on the fundamental diagram we refer to Zhang (2012). The measurement methods usually rely on a discretization scheme chosen arbitrarily, in both space and time (Seyfried et al., 2010; Daamen and Hoogendoorn, 2003). This may generate noise in the data, and the results may be highly sensitive to minor changes of discretization (Openshaw, 1983; Liddle et al., 2011). In order to be as much independent from the aggregation level as possible, we rely on a data-driven measurement method inspired by the one proposed by Steffen and Seyfried (2010). This method is based on the

spatial discretization that is adjusted to the data itself through the use of Voronoi diagrams (Okabe et al., 2000).

The Voronoi space decomposition (Okabe et al., 2000) assigns a personal region $V_i(t)$ to each pedestrian i , in such a way that each point in the personal region is closer to i than to any other pedestrian, with respect of the Euclidean distance

$$V_i(t) = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} \right\|_2 \leq \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_j(t) \\ y_j(t) \end{pmatrix} \right\|_2, \forall j \right\}. \quad (3)$$

In the presence of sampled data defined by (2), we have, for each $s = 0, \dots, f$ and each pedestrian i

$$V_{is} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_{is} \\ y_{is} \end{pmatrix} \right\|_2 \leq \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_{js} \\ y_{js} \end{pmatrix} \right\|_2, \forall j \right\}. \quad (4)$$

We assume that each point (x, y) in space is associated with a unique Voronoi region at time t_s , corresponding to the region associated with pedestrian i , that is $V(x, y, t_s) = V_{is}$. Note that if (x, y) is exactly on the border between two or more regions, the unique region associated to it has to be arbitrarily defined.

Given the space discretization specified above, the density of pedestrians at position (x, y) at time t_s is

$$k(x, y, t_s) = \frac{1}{|V(x, y, t_s)|}, \quad (5)$$

where $V(x, y, t_s)$ is the unique Voronoi region that contains (x, y) at time t_s , and $|V(x, y, t_s)|$ is the area of $V(x, y, t_s)$. The unit is the number of pedestrians per surface unit (typically, square meter).

We refer to Steffen and Seyfried (2010) and Nikolić et al. (2014) for detailed discussions of this approach.

The velocity of pedestrian i at time t is given by

$$\vec{v}_i(t) = v_i(t) \vec{d}_i(t), \quad (6)$$

where $\vec{d}_i(t)$ is the (normalized) direction of pedestrian i at time t and $v_i(t)$ is the magnitude of the velocity vector, or speed. If the functions $x_i(t)$ and $y_i(t)$ in (1) are differentiable in t , it is defined as

$$v_i(t) = \sqrt{\left(\frac{dx_i(t)}{dt} \right)^2 + \left(\frac{dy_i(t)}{dt} \right)^2}. \quad (7)$$

Note that this definition assumes that the direction of the flow is unique at each point in time and space. It may therefore not be appropriate for the analysis of multidirectional flow. In the presence of discretized data, the speed is approximated using finite differences, that is

$$v_{is} = \sqrt{\left(\frac{\Delta x_{is}}{\Delta t}\right)^2 + \left(\frac{\Delta y_{is}}{\Delta t}\right)^2}, \quad (8)$$

where $\Delta x_{is} = x_{i,s+1} - x_{i,s-1}$, $\Delta y_{is} = y_{i,s+1} - y_{i,s-1}$, and $\Delta t = t_{s+1} - t_{s-1}$.

5 Empirical analysis

The speed-density profiles corresponding to the Lausanne and the Delft case studies are obtained from the measurement method presented in Section 4. In Figure 3, each circle corresponds to one observation, that is, one pedestrian at one specific time in the horizon. The x coordinate of the circle corresponds to the density, calculated from (5), and its y coordinate corresponds to the speed calculated from (8).

Figure 3a plots 270,291 observations corresponding to the peak hour of February 12, 2013 for the Lausanne case study. The same pattern was observed on any weekday. Figure 3b plots 119,156 observations for the Delft case study.

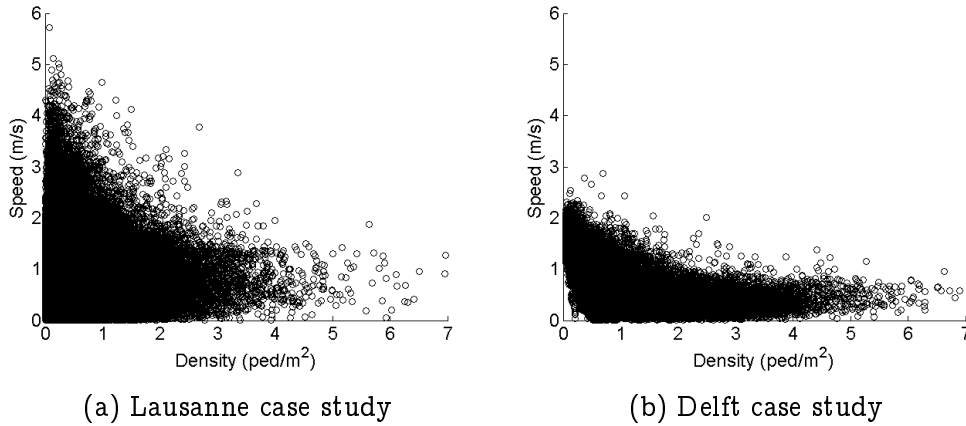


Figure 3: Speed-density profiles

A high scattering is observed in both cases (Figure 3). The density ranges from 0 to approximately 7 pedestrians per square meter. In the

Lausanne case, the speed ranges from 0 to 5.72 meters per second (that is about 21 km/h), and 99% of the observations are between 0 and 2.42 meters per second (that is about 9 km/h). In the Delft case, the speed ranges from 0 to 2.87 meters per second (that is about 10 km/h). The difference in the speed distribution is attributed to the controlled nature of the experiment in Delft, where individuals were instructed to walk at normal speed, resulting to a lower variance compared to Lausanne, where no instruction was given. For the same reason, low speeds were not observed at low density in Delft, contrarily to Lausanne.

To investigate this data in more details, the speed distributions at various density levels are presented in Figures 4 and 5 for the Lausanne and Delft case study, respectively. In both cases, a higher level of variability is noticeable at lower densities, compared to higher densities where the distribution of speed is less spread and shifted towards lower values.

The deterministic models for the speed-density relationship proposed in the literature (Section 2.2) appear to be inadequate for representing the observed patterns. Clearly, density is not the only factor influencing pedestrians' speed.

In order to take the nature of the data into account and to characterize the observed phenomena, we propose next a probabilistic model for the speed-density relationship.

6 Modeling framework

We assume that there are two subpopulations of pedestrians, associated with two different (unspecified) types of behavior: fast and slow pedestrians. These subpopulations are further assumed to be characterized by corresponding components in the overall speed distribution. This assumption is in line with the pattern observed and discussed in Section 5 and will be validated in Section 7 and Section 8.

6.1 Model specification

Consistently with the assumption of two subpopulations, the probabilistic speed-density relationship is specified by using a mixture of models. The velocities are modeled as random variables. We denote by $f_{\text{slow}}(\xi|k; \theta_{\text{slow}})$

the probability density function of the speed of slow pedestrians, conditional to the density k , and by $f_{\text{fast}}(\xi|k; \theta_{\text{fast}})$ the probability density function of the speed of fast pedestrians. We consider a pedestrian to be “fast” when her speed is beyond a threshold denoted $v_m(k)$, that varies with the density level. We assume that it is a random variable with a symmetric triangular distribution defined on the interval $[\bar{v}_m(k) - \sigma, \bar{v}_m(k) + \sigma]$, where $\bar{v}_m(k)$ is a parameter representing the mean, and $0 \leq \sigma \leq \bar{v}_m(k)$

$$f_{v_m(k)}(\xi; \bar{v}_m(k), \sigma) = \begin{cases} \frac{\xi - \bar{v}_m(k) + \sigma}{\sigma^2}, & \bar{v}_m(k) - \sigma \leq \xi \leq \bar{v}_m(k) \\ \frac{\bar{v}_m(k) + \sigma - \xi}{\sigma^2}, & \bar{v}_m(k) < \xi \leq \bar{v}_m(k) + \sigma \\ 0, & \xi < \bar{v}_m(k) - \sigma \text{ or } \xi > \bar{v}_m(k) + \sigma. \end{cases} \quad (9)$$

The specification of $\bar{v}_m(k)$ is typically a deterministic speed-density relationship, as those listed in Table 2. For instance, in the case studies presented below, we have adopted

$$\bar{v}_m(k) = v_f - k\gamma, \quad (10)$$

which is inspired by Older (1968), Navin and Wheeler (1969), Fruin (1971), Tanaboriboon et al. (1986) and Lam et al. (1995), and

$$\bar{v}_m(k) = v_f \exp\left(-\left(\frac{k}{\theta}\right)^\gamma\right), \quad (11)$$

inspired by Tregenza (1976). But any other relevant specification can be accommodated by the framework.

The probability density function of the speed for the entire population is then defined as

$$f_v(\xi|v_m(k), \alpha_k, \beta_k, \lambda) = f_{\text{slow}}(\xi|v_m(k), \alpha_k, \beta_k) \Pr(\xi \leq v_m(k)) + f_{\text{fast}}(\xi|v_m(k), \beta_k, \lambda) \Pr(\xi \geq v_m(k)), \quad (12)$$

where $\Pr(\xi \leq v_m(k))$ and $\Pr(\xi \geq v_m(k))$ are the probabilities that we are dealing with a slow, respectively fast, pedestrian. We adopt a simple specification for the parameters α_k , β_k . We assume that they depend on the density k in the following way

$$\alpha_k(a_\alpha, b_\alpha) = a_\alpha k + b_\alpha, \quad (13)$$

and

$$\beta_k(a_\beta, b_\beta) = a_\beta k + b_\beta. \quad (14)$$

The assumptions regarding the distribution of the mixture components are based on the pattern exhibited by the data itself (Section 5). The empirical distribution of the speed values with respect to density (Figure 4 and Figure 5) suggests the following models. We propose a linear model for the distribution of the speed of slow pedestrians

$$f_{\text{slow}}(\xi|k) = \frac{\beta_k - \alpha_k}{v_m(k)} \xi + \alpha_k, \quad (15)$$

where $v_m(k) \geq 0$ is the transition speed at the level of density k , that is the speed threshold between being “slow” and “fast”, and $\alpha_k \geq 0$ and $\beta_k \geq 0$ are parameters depending on k . They are such that $f_{\text{slow}}(0|k) = \alpha_k$ and $f_{\text{slow}}(v_m(k)|k) = \beta_k$. We propose an exponential model for the speed of fast pedestrians

$$f_{\text{fast}}(\xi|k) = \exp(-\lambda\xi + \log(\beta_k) + \lambda v_m(k)), \quad (16)$$

where $v_m(k)$ and β_k are defined as above, and $\lambda \geq 0$ is a additional parameter, defining the rate of the exponential distribution. Note that, if $\xi = v_m(k)$, the two values coincide, and are both equal to β_k .

The parameters characterizing the distribution of each component are illustrated in Figure 6.

Putting everything together, the probability density function of the speed is

$$f_v(\xi|k; \alpha_k(a_\alpha, b_\alpha), \beta_k(a_\beta, b_\beta), \lambda, \bar{v}_m(k), \sigma) = \int_{t=0}^{\infty} f_v(\xi|t; \alpha_k(a_\alpha, b_\alpha), \beta_k(a_\beta, b_\beta), \lambda) f_{v_m(k)}(t; \bar{v}_m(k), \sigma) dt, \quad (17)$$

where $f_{v_m(k)}(t; \bar{v}_m(k), \sigma)$ is defined by (9) and $f_v(\xi|t; \alpha_k(a_\alpha, b_\alpha), \beta_k(a_\beta, b_\beta), \lambda)$ is defined by (12).

The parameters a_α , b_α , a_β , b_β , λ , $\bar{v}_m(k)$, and σ are parameters to be estimated, for instance by maximum likelihood estimation. In the following, the model is called PedMixFD, for **P**edestrian **M**ixture **F**undamental **D**igram.

7 Case study: Lausanne

We illustrate and validate now the model on the Lausanne dataset, introduced in Section 3.1.

The dataset used for estimation consists of 1,269,393 pairwise speed-density observations corresponding to the peak hour of February 12, 13, 14, 15 and 18, 2013.

The descriptive statistics of the estimation dataset are presented in Table 3. The dataset is categorized according to six levels of service (LOS) proposed by Fruin (1971) for pedestrian facilities, labeled from A to F. Table 3 shows that the largest part of the observations falls below the LOS F. Actually, 99% of the observations are below 2.06 ped/m².

Level Of Service	Number of observations
A ($k \leq 0.31$ ped/m ²)	644546
B ($k \in (0.31 - 0.43$ ped/m ²])	174116
C ($k \in (0.43 - 0.71$ ped/m ²])	229808
D ($k \in (0.71 - 1.11$ ped/m ²])	133812
E ($k \in (1.11 - 2.17$ ped/m ²])	76725
F ($k > 2.17$ ped/m ²)	10386

Table 3: Estimation data classified according to LOS (Fruin, 1971) - Lausanne case study

The estimation results for the model presented in the previous section are shown in Table 4. The parameter $\bar{v}_m(k)$ is specified in the model (10) by Older (1968), Navin and Wheeler (1969), Fruin (1971), Tanaboriboon et al. (1986) and Lam et al. (1995). All estimates have the expected sign and value, indicating the good model specification. The results also show the low standard errors of all parameters. Note that the panel nature of the data (Baltagi, 2008) has been ignored here.

The positive sign of the parameter α_α shows that α_k , that is the likelihood of low speeds, increases with density. Similarly, the positive sign of the parameter α_β shows that β_k , that is the likelihood of the mode of the speed distribution, also increases with density.

The signs and the estimated values of the parameters v_f and γ are consistent with the ones reported in the literature (see Section 2) and with the trend observed in the data.

Note that various specifications of the model have been investigated, but not all of them are reported in this paper. For instance, we tried for \bar{v}_m the specification proposed by Weidmann (1993) (Table 2) but obtained a poorer fit. Indeed, the Bayesian information criterion - BIC (Wasserman, 2000)

Parameter	Value	Std err
a_α	0.0393	$7.57e^{-07}$
b_α	0.00699	$1.07e^{-06}$
a_β	0.00490	$7.96e^{-07}$
b_β	0.142	$1.75e^{-06}$
λ	3.53	$1.12e^{-06}$
v_f	1.29	$2.40e^{-06}$
γ	0.0512	$5.17e^{-07}$
σ	0.0383	$3.77e^{-06}$
$\log \mathcal{L}$	-783933.016	
Number of parameters	8	
Number of observations	1269393	

Table 4: Estimation results - Lausanne case study

was 1690285.550, as opposed to 1567978.464 for PedMixFD.

7.1 Comparison with deterministic models

The performance of the proposed probabilistic model at the aggregate level is compared with the deterministic models proposed in the literature (Table 2). We have estimated the parameters of these models using linear regression on our dataset.

For PedMixFD, the average speed is given by

$$\bar{v}_{\text{PedMixFD}}(k) = \int_0^\infty \xi f_v(\xi|k; \alpha_k(a_\alpha, b_\alpha), \beta_k(a_\beta, b_\beta), \lambda, \bar{v}_m(k), \sigma) d\xi. \quad (18)$$

In (18) $f_v(\xi|k; \alpha_k(a_\alpha, b_\alpha), \beta_k(a_\beta, b_\beta), \lambda, \bar{v}_m(k), \sigma)$ refers to mixture distribution given by (17), with the parameters described in Table 4.

Figure 7 shows a comparison among four deterministic models, the aggregate speed calculated from PedMixFD, and the observed values. The analysis is performed for density levels ranging from 0 to 2.06 ped/m², corresponding to 99% of the observed values. The goodness of fit is reported in Table 5, where

$$\text{MSE} = \frac{1}{m} \sum_{\ell=1}^m (\bar{v}_{\text{model}}(k_\ell) - \bar{v}_{\text{data}}(k_\ell))^2, \quad (19)$$

where $m = 22$ and $k_\ell = 2.06(\ell - 1)/(m - 1)$. The value $\bar{v}_{\text{data}}(k_\ell)$ is the average of the observed speeds corresponding to densities ranging from $(k_\ell + k_{\ell-1})/2$ and $(k_\ell + k_{\ell+1})/2$. Although the PedMixFD has been calibrated at the disaggregate level, it is interesting to observe that it achieves the best fit at the aggregate level.

Model	Weidmann (1993)	Tregenza (1976)	Rastogi et al. (2013)	Linear	PedMixFD
MSE	$6.69e^{-03}$	$1.82e^{-03}$	$1.90e^{-03}$	$1.87e^{-03}$	$1.27e^{-03}$

Table 5: Goodness of Fit (MSE) - Lausanne case study

7.2 Kolmogorov-Smirnov validation

The validation is performed by comparing the distribution functions of the estimated model, and the empirical distributions from the dataset. The analysis is carried out at different density levels. Figure 8 shows the probability density functions of the model (model pdf) and the data (empirical histogram) at the same density level. The corresponding cumulative density functions (model cdf and empirical cdf) are plotted in Figure 9. Qualitatively, the match between the two is pretty satisfactory.

For the quantitative analysis, we use the Kolmogorov-Smirnov probability distance metric (Massey, 1951)

$$D_k = \max_v |F_{\text{model}}(v|k) - F_{\text{data}}(v|k)|, \quad (20)$$

where $F_{\text{model}}(v|k)$ corresponds to the model cdf and $F_{\text{data}}(v|k)$ to the empirical cdf. This metric represents the maximum value of the absolute vertical difference between the two cumulative distribution functions. It is reported in Figure 10a.

We calculate the p-value of the Kolmogorov-Smirnov statistic using simulation (Ross, 2013a, p. 257), with 50 simulation runs. The model (17) is simulated using the rejection method (Ross, 2013b, Section 5.2) on draws from a Rayleigh distribution. The results are shown in Figure 10b. They suggest that there is no evidence in the data to reject PedMixFD at significance level 0.05 for all levels of density except maybe for the one corresponding to densities close to zero.

7.3 Specification test

In order to test the robustness of the proposed specification, we have performed the validation that consists in splitting the dataset into two subsets. The model is re-estimated on one subset and the remaining data, unused for estimation, is used for validation purposes. The procedure consisting of the following steps is repeated 100 times:

1. A sample of 80% of pairwise speed-density observations is selected using simple random selection.
2. The parameters of the model are estimated using the generated sample.
3. The Kolmogorov-Smirnov statistic D_k (20) is calculated to compare the estimated model and the data on the remaining 20% of the dataset.

In Figure 11a, we compare the value of D_k calculated on the full dataset (in dashed line) with the values calculated with the above mentioned procedure. The 100 values are summarized using a box plot at each level of density. These results are satisfactory. The specification is robust and no overfitting is detected.

To be more precise, we also calculate the p-value for each value of D_k calculated with the above mentioned procedure. For this purpose we use simulation (Ross, 2013a, p. 252), with 100 simulation runs. The box plot of the estimated p-values are shown in Figure 11b. The results do not allow to reject the hypothesis that the data and the model follow the same distribution, at a usual level of significance.

8 Case study: Delft

In this section we illustrate and validate the model on the Delft dataset, described in Section 3.2.

The dataset used for estimation consist of 119,156 pairwise speed-density values observed upstream of the bottleneck. The data has been classified according to the LOS standard of Fruin (1971), showing that now the majority of the observations corresponds to the LOS C, D, E and F (Table 6). Consequently 99% of the data is below the density value of 3.9 ped/m².

This is as expected, given the existence of flow constraint (in the form of a narrow bottleneck) that in this case causes congestion upstream of the bottleneck.

Level Of Service	Number of observations
A ($k \leq 0.31$ ped/m ²)	9288
B ($k \in (0.31 - 0.43$ ped/m ²])	6967
C ($k \in (0.43 - 0.71$ ped/m ²])	20497
D ($k \in (0.71 - 1.11$ ped/m ²])	21540
E ($k \in (1.11 - 2.17$ ped/m ²])	37114
F ($k > 2.17$ ped/m ²)	23750

Table 6: Estimation data classified according to LOS (Fruin, 1971) - Delft case study

The parameters of the model (17) have been estimated, where $\bar{v}_m(k)$ is specified by the model (11) by Tregenza (1976). The estimation results are shown in Table 7. The sign and the magnitude of the parameters are as expected. The results also indicate a high significance of the estimated values.

Parameter	Value	Std err
a_α	0.164	$1.47e^{-04}$
b_α	0.244	$1.96e^{-04}$
a_β	0.166	$2.17e^{-04}$
b_β	0.965	$1.62e^{-05}$
λ	6.89	$1.13e^{-05}$
v_f	1.87	$1.20e^{-04}$
θ	1.13	$2.93e^{-04}$
γ	0.545	$8.62e^{-05}$
σ	0.0357	$9.77e^{-05}$
$\log \mathcal{L}$	-3768.931	
Number of parameters	9	
Number of observations	119156	

Table 7: Estimation results - Delft case study

The positive sign of the parameter a_α shows that α_k , that is the likelihood of low speeds, increases with density. Similarly, the positive sign of

the parameter α_β shows that β_k , that is the likelihood of the mode of the speed distribution, also increases with density.

The signs and the estimated values of the parameters v_f , θ and γ of the model inspired by Tregenza (1976) are in accordance with the trend observed in the data.

The value of the parameter λ is higher than the one for the Lausanne case study, which is consistent with the reduced range of speed values in the data. Finally, the value of the parameter σ indicates that variability of the mode of the speed distribution is slightly lower than that of Lausanne case study.

8.1 Comparison with deterministic models

The performance of the proposed model is compared with the deterministic models proposed in the literature (Table 2), using the same procedure as for the Lausanne case study. The models are compared in Figure 12, and the goodness of fit measures are reported in Table 8. Again, PedMixFD exhibits the best fit.

Model	Weidmann (1993)	Tregenza (1976)	Rastogi et al. (2013)	Linear	PedMixFD
MSE	$3.79e^{-02}$	$1.93e^{-02}$	$3.37e^{-02}$	$4.56e^{-02}$	$1.69e^{-02}$

Table 8: Goodness of Fit (MSE) - Delft case study

8.2 Kolmogorov-Smirnov validation

The agreement between the model predictions and the observations from the estimation dataset is illustrated in Figure 13 and Figure 14. The Kolmogorov-Smirnov distances (20) between the model cdf and the empirical cdf are illustrated in the Figure 15a.

The agreement between the model predictions with data appears to be less satisfactory for lower density levels, as we have fewer data with low speed at low density levels. This is an artefact of the experimental nature of the data (see discussion in Section 5). The quality of the fit for higher density levels (which is of greater interest for applications anyway) is satisfactory. Figure 15a shows that the smallest Kolmogorov-Smirnov distances correspond to the density levels which are characterized by the largest number of observations (Table 6).

The p-values are estimated using the procedure described in Section 7.2 and shown in Figure 15b. Again, there is no evidence in the data to reject PedMixFD at significance level 0.05 for most of the density levels. The p-values less than 0.05 are observed for lower density levels, up to 0.3 ped/m², and for density levels greater than 3.5 ped/m². In the former case low p-values are caused by the experimental nature of the data (as discussed above), while in the latter a low number of observations (0.02% of the data) is insufficient to reach any conclusion.

8.3 Specification test

We test the robustness of the model specification by performing the validation using 80% of the data for estimation and the remaining 20% for validation (see Section 7.3). The Kolmogorov-Smirnov statistics for different density levels from 100 simulation runs and corresponding p-values are shown using box plot representation in Figure 16a and Figure 16b, respectively. The above results validate the model also for the Delft case study.

9 Conclusion and future work

In this paper a novel speed-density relationship for pedestrian traffic is proposed. Different from the deterministic approaches in the literature, it is a probabilistic model designed to account for the heterogeneity of speed at a given density level, as observed in the data.

Various tests on two different case studies validate the specification of the model. Moreover, the model is shown to outperform traditional deterministic models at the aggregate level as well.

The presented work has both theoretical and practical implications. It can be combined with a conservation principle in dynamic continuum and discrete models (Hughes, 2002; Hänseler et al., 2014), leading to probabilistic conservation laws for the representation of the pedestrian dynamics. This would allow the detailed analysis of the effects of heterogeneity on pedestrian flows. The suggested model may be utilized as such by practitioners for the evaluation and optimization of the level of service of pedestrian facilities. Contrasted with existing approaches, it yields a more realistic representation of the empirically observed phenomena.

Nonetheless, some aspects require further investigation. First, the panel nature of the data should be exploited. Second, the described framework as such is insufficient to explain the multi-directional nature of pedestrian flows. As further steps we will explore the possibility of addressing this issue by adapting the existing definitions of pedestrian traffic characteristics through a stream-based approach and a data-driven spatio-temporal discretization framework (Nikolić and Bierlaire, 2014).

Finally, we plan to combine this data driven approach with a behavioral approach, where the heterogeneity of speeds is explicitly explained, using variables such as trip purpose, age, walking in group, walking with luggages, etc.

Acknowledgements

This research is supported by the Swiss National Science Foundation Grant 200021-141099 "Pedestrian dynamics: flows and behavior". The authors would like to thank SBB-CFF-FFS , who provided us with the dataset collected in the Lausanne train station in the framework of the project "Ped-Flux: Pedestrian flow modeling in train stations", and Winnie Daamen and Serge Hoogendoorn for the data from controlled experiments performed at the Technical University of Delft in the Netherlands. We are also thankful to our group members Riccardo Scarinci, Shadi Sharif Azadeh, Flurin Hänseler and Matthieu de Lapparent for their invaluable assistance and suggestions.

References

- Alahi, A., Bierlaire, M. and Vanderghelynst, P. (2014). Robust real-time pedestrians detection in urban environments with a network of low resolution cameras, *Transportation Research Part C: Emerging Technologies* **39**: 113–128.
- Alahi, A., Jacques, L., Boursier, Y. and Vanderghelynst, P. (2011). Sparsity driven people localization with a heterogeneous network of cameras, *Journal of Mathematical Imaging and Vision* **41**(1-2): 39–58.
- Baltagi, B. (2008). *Econometric analysis of panel data*, Vol. 1, John Wiley & Sons.
- Bauer, D., Brandle, N., Seer, S., Ray, M. and Kitazawa, K. (2009). Measurement of pedestrian movements: A comparative study on various existing systems, *Pedestrian Behavior: Models, Data Collection and Applications*, Emerald Group Publishing Limited: Bingley, UK.
- Bowman, B. L. and Vecellio, R. L. (1994). Pedestrian walking speeds and conflicts at urban median locations, *Transportation research record* (1438): 67–73.
- Cheah, J. Y. and Smith, J. M. (1994). Generalized M/G/c/c state dependent queueing models and pedestrian traffic flows, *Queueing Systems* **15**(1-4): 365–386.
- Cheung, C. and Lam, W. H. (1998). Pedestrian route choices between escalator and stairway in mtr stations, *Journal of transportation engineering* **124**(3): 277–285.
- Daamen, W. (2004). *Modelling passenger flows in public transport facilities*, PhD thesis, Delft University of Technology, Delft.
- Daamen, W. and Hoogendoorn, S. P. (2003). Controlled experiments to derive walking behaviour, *European Journal of Transport and Infrastructure Research* **3**(1): 39–59.
- Daamen, W., Hoogendoorn, S. P. and Bovy, P. H. (2005). First-order pedestrian traffic flow theory, *Transportation Research Record: Journal of the Transportation Research Board* **1934**(1): 43–52.

- DiNenno, P. J. (2002). *SFPE handbook of fire protection engineering*, National Fire Protection Association, Quincy, Massachusetts.
- Drake, J., Schofer, J. and May, A. (1967). A statistical analysis of speed-density hypotheses, *Highway Research Record* **154**: 53–87.
- Duives, D., Daamen, W. and Hoogendoorn, S. (2014). Anticipation behavior upstream of a bottleneck, *Transportation Research Procedia* **2**: 43–50.
- Fruin, J. J. (1971). Designing for pedestrians: A level-of-service concept, number 355, Highway Research Board, Washington, DC, pp. 1–15.
- Greenshields, B., Bibbins, J., Channing, W. and Miller, H. (1935). A study of traffic capacity, *Proceedings of the Highway Research Board*, Vol. 14, Highway Research Board, Washington, DC.
- Hänseler, F., Bierlaire, M., Farooq, B. and Mühlematter, T. (2014). A macroscopic loading model for time-varying pedestrian flows in public walking areas, *Transportation Research Part B: Methodological* **69**: 60 – 80.
- Hughes, R. L. (2002). A continuum theory for the flow of pedestrians, *Transportation Research Part B: Methodological* **36**(6): 507–535.
- Jabari, S. E., Zheng, J. and Liu, H. X. (2014). A probabilistic stationary speed–density relation based on Newell’s simplified car-following model, *Transportation Research Part B: Methodological* **68**: 205–223.
- Lam, W. H., Morrall, J. F. and Ho, H. (1995). Pedestrian flow characteristics in Hong Kong, *Transportation Research Record* (1487): 56–62.
- Liddle, J., Seyfried, A., Steffen, B., Klingsch, W., Rupprecht, T., Winkens, A. and Boltes, M. (2011). Microscopic insights into pedestrian motion through a bottleneck, resolving spatial and temporal variations, *arXiv:1105.1532*.
- Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit, *Journal of the American statistical Association* **46**(253): 68–78.

- Navin, F. and Wheeler, R. (1969). Pedestrian flow characteristics, *Traffic Engineering, Inst Traffic Engr* **39**.
- Newell, G. F. (1961). Nonlinear effects in the dynamics of car following, *Operations Research* **9**(2): 209–229.
- Nikolić, M. and Bierlaire, M. (2014). Pedestrian-oriented flow characterization, *Transportation Research Procedia* **2**: 359–366.
- Nikolić, M., Bierlaire, M. and Farooq, B. (2014). Probabilistic speed-density relationship for pedestrians based on data driven space and time representation, *Proceedings of the Swiss Transportation Research Conference*, Ascona, Switzerland.
- Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N. (2000). *Spatial tessellations: concepts and applications of Voronoi diagrams*, Second edn, Wiley, New York.
- Older, S. (1968). Movement of pedestrians on footways in shopping streets, *Traffic engineering and control* **10**: 160–163.
- Openshaw, S. (1983). The modifiable areal unit problem, *Concepts and Techniques in Modern Geography*, number 38, Geobooks, Norwich, England.
- Rastogi, R., Ilango, T. and Chandra, S. (2013). Pedestrian flow characteristics for different pedestrian facilities and situations, *European Transport* **53**.
- Ross, S. (2013a). Chapter 11 - Statistical Validation Techniques, *Simulation (Fifth Edition)*, Academic Press, pp. 247 – 270.
- Ross, S. (2013b). Chapter 5 - Generating Continuous Random Variables, *Simulation (Fifth Edition)*, Academic Press, pp. 69 – 96.
- Seyfried, A., Boltes, M., Kähler, J., Klingsch, W., Portz, A., Rupprecht, T., Schadschneider, A., Steffen, B. and Winkens, A. (2010). Enhanced empirical data for the fundamental diagram and the flow through bottlenecks, *Pedestrian and Evacuation Dynamics 2008*, Springer.

- Steffen, B. and Seyfried, A. (2010). Methods for measuring pedestrian density, flow, speed and direction with minimal scatter, *Physica A: Statistical mechanics and its applications* **389**(9): 1902–1910.
- Tanaboriboon, Y., Hwa, S. S. and Chor, C. H. (1986). Pedestrian characteristics study in singapore, *Journal of Transportation Engineering* **112**(3): 229–235.
- Tregenza, P. (1976). *The design of interior circulation*, Van Nostrand Reinhold, New York.
- Underwood, R. T. (1961). Speed, volume, and density relationships: Quality and theory of traffic flow, *Yale Bureau of Highway Traffic, New Haven*.
- van Wageningen-Kessels, F. (2013). *Multi-class continuum traffic flow models: analysis and simulation methods*, PhD thesis, Delft University of Technology/TRAIL Research school, Delft.
- Wang, H., Li, J., Chen, Q.-Y. and Ni, D. (2009). Speed-density relationship: From deterministic to stochastic, *Transportation Research Board 88th Annual Meeting, Washington, DC*.
- Wasserman, L. (2000). Bayesian model selection and model averaging, *Journal of mathematical psychology* **44**(1): 92–107.
- Weidmann, U. (1993). Transporttechnik der fussgänger, *Technical Report Schriftenreihe des IVT Nr. 90*, Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau, ETH Zürich. (In German).
- Zhang, J. (2012). *Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries*, PhD thesis, Forschungszentrum Jülich.

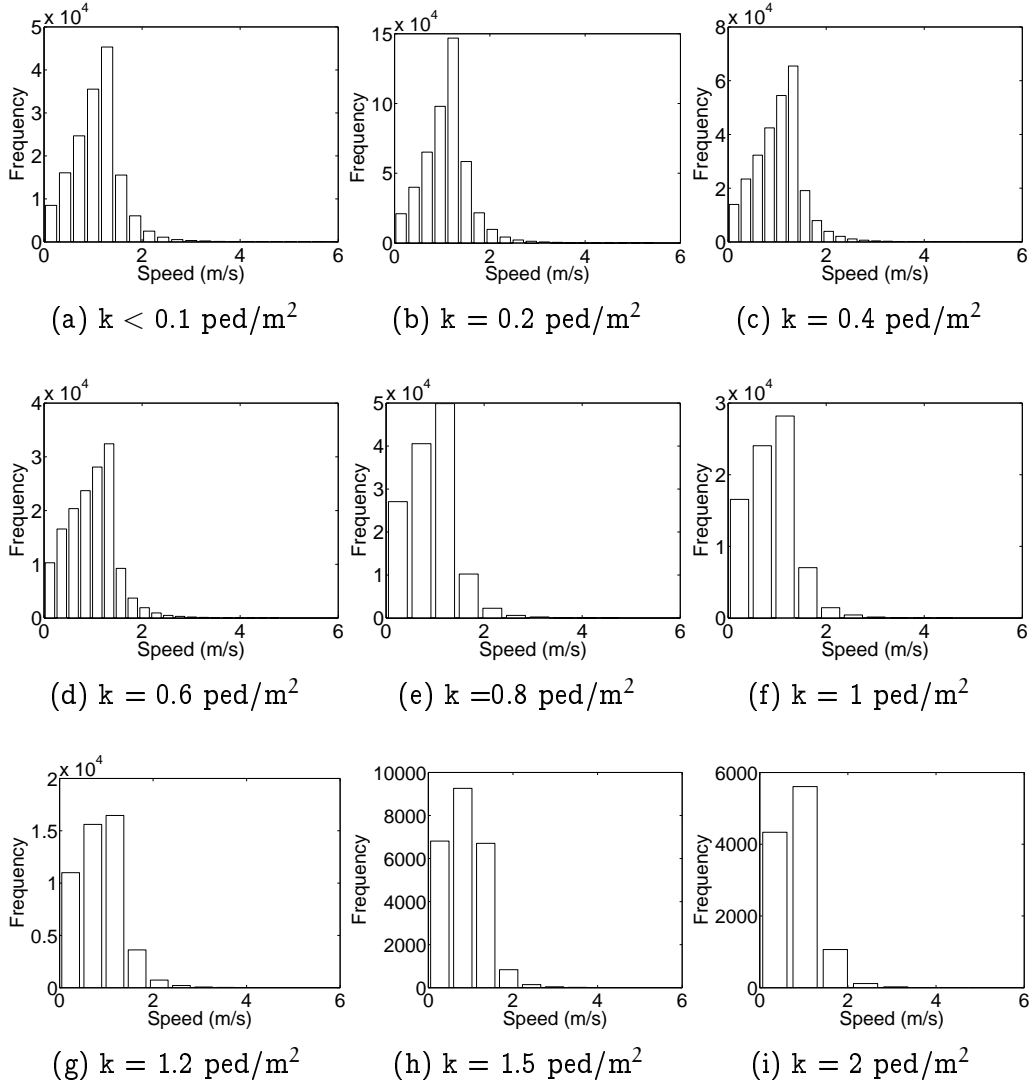


Figure 4: Speed distributions for different density levels - Lausanne case study

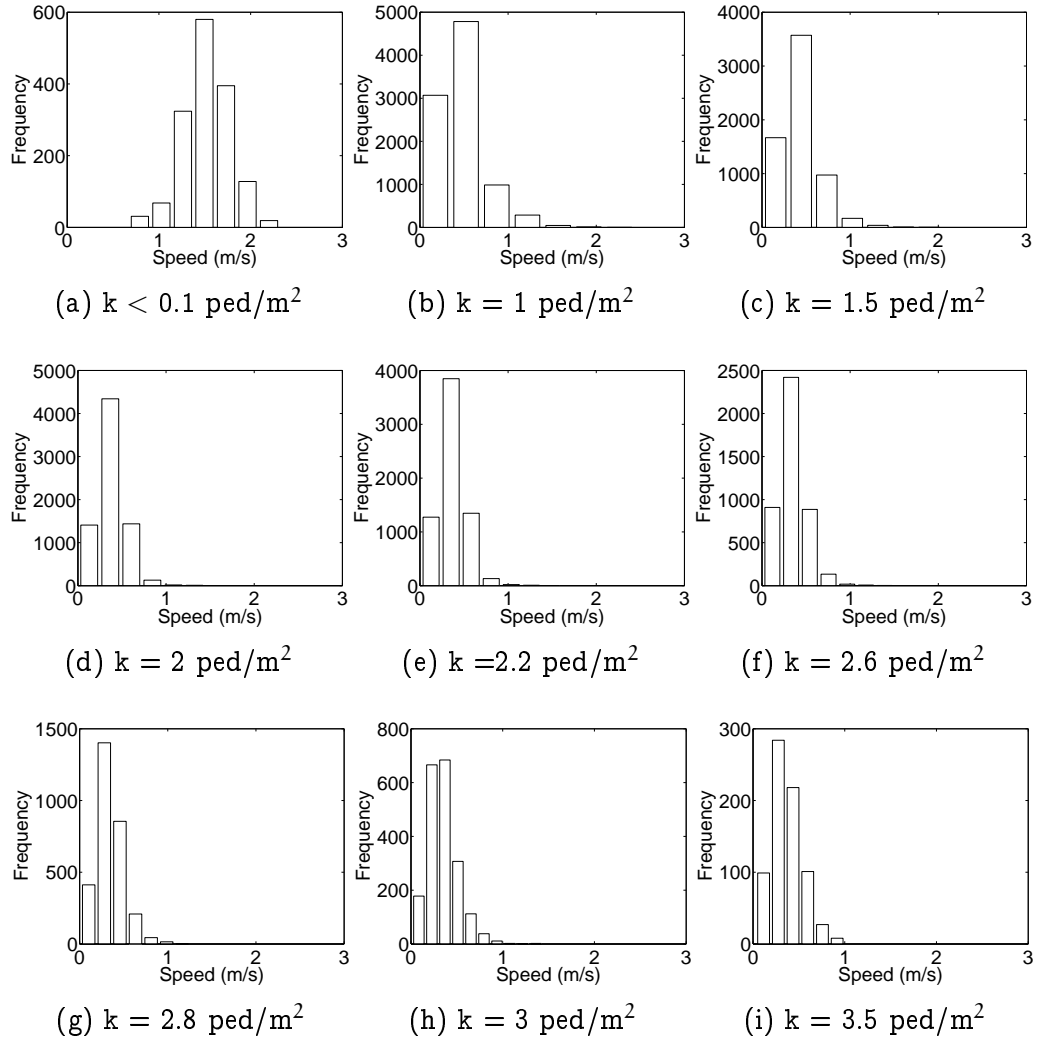


Figure 5: Speed distributions for different density levels - Delft case study

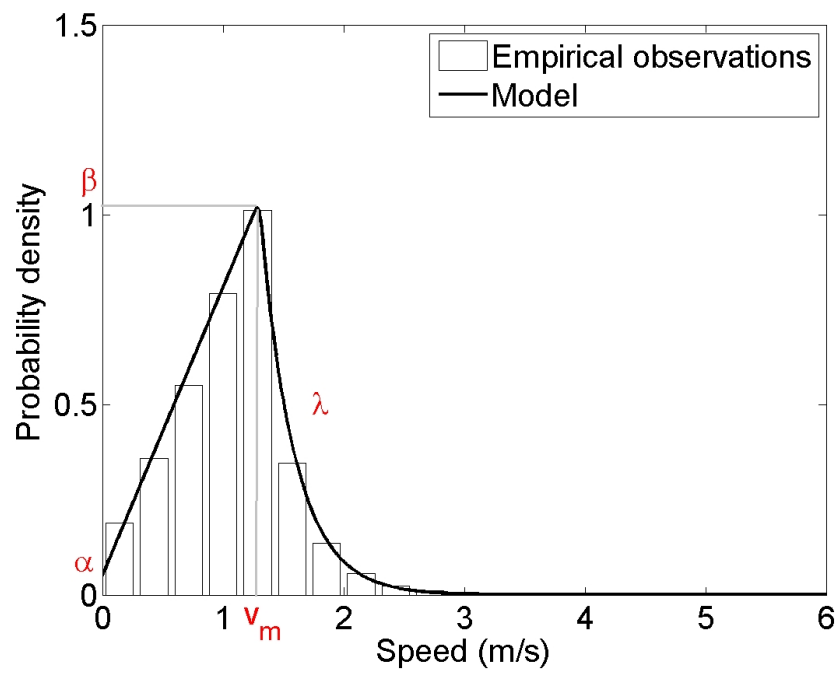


Figure 6: Illustration of the model - one density level

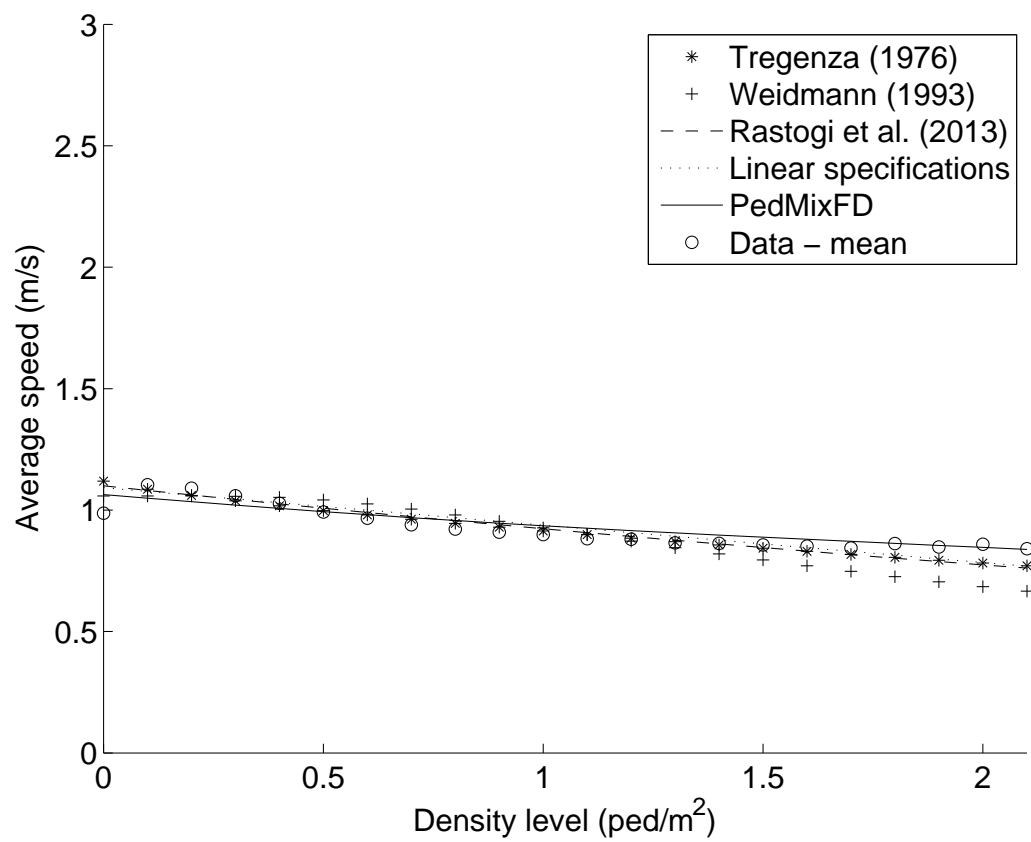


Figure 7: Comparison between deterministic models predictions and the aggregated probabilistic model predictions - Lausanne case study

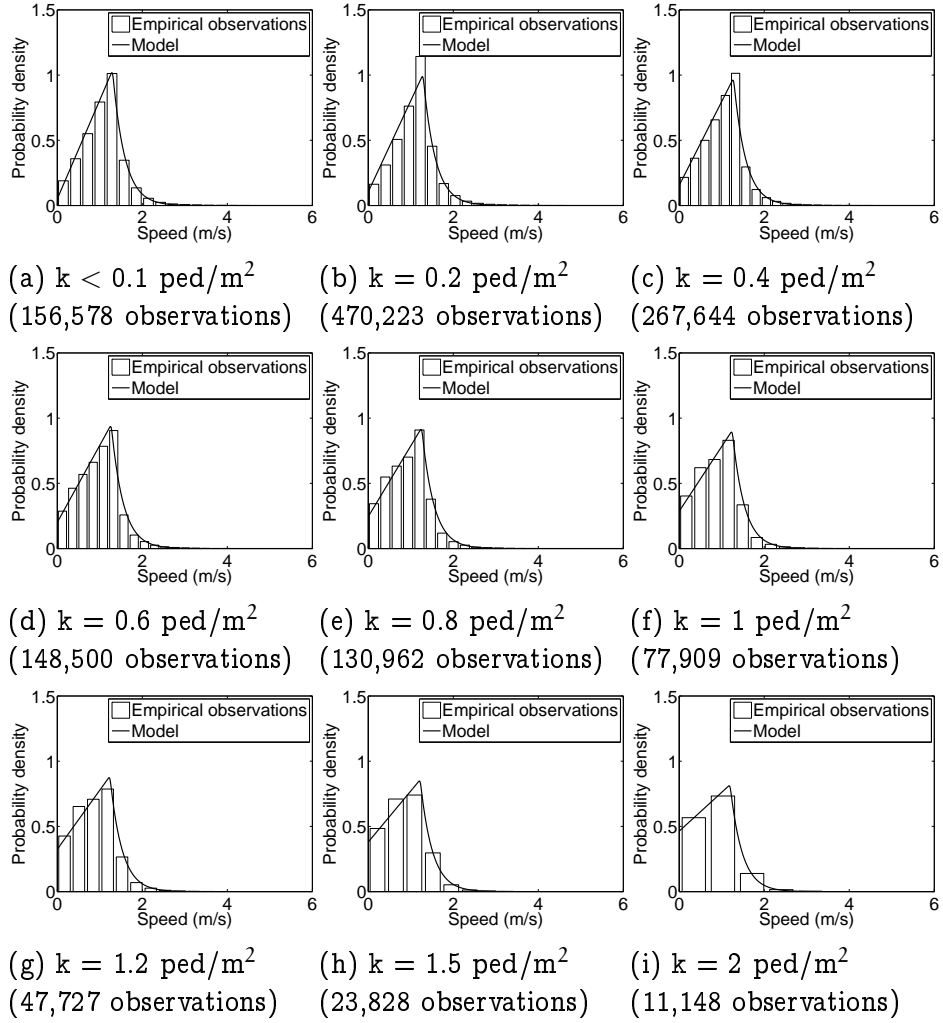


Figure 8: Comparison between model predictions (probability density) and empirical observations - Lausanne case study

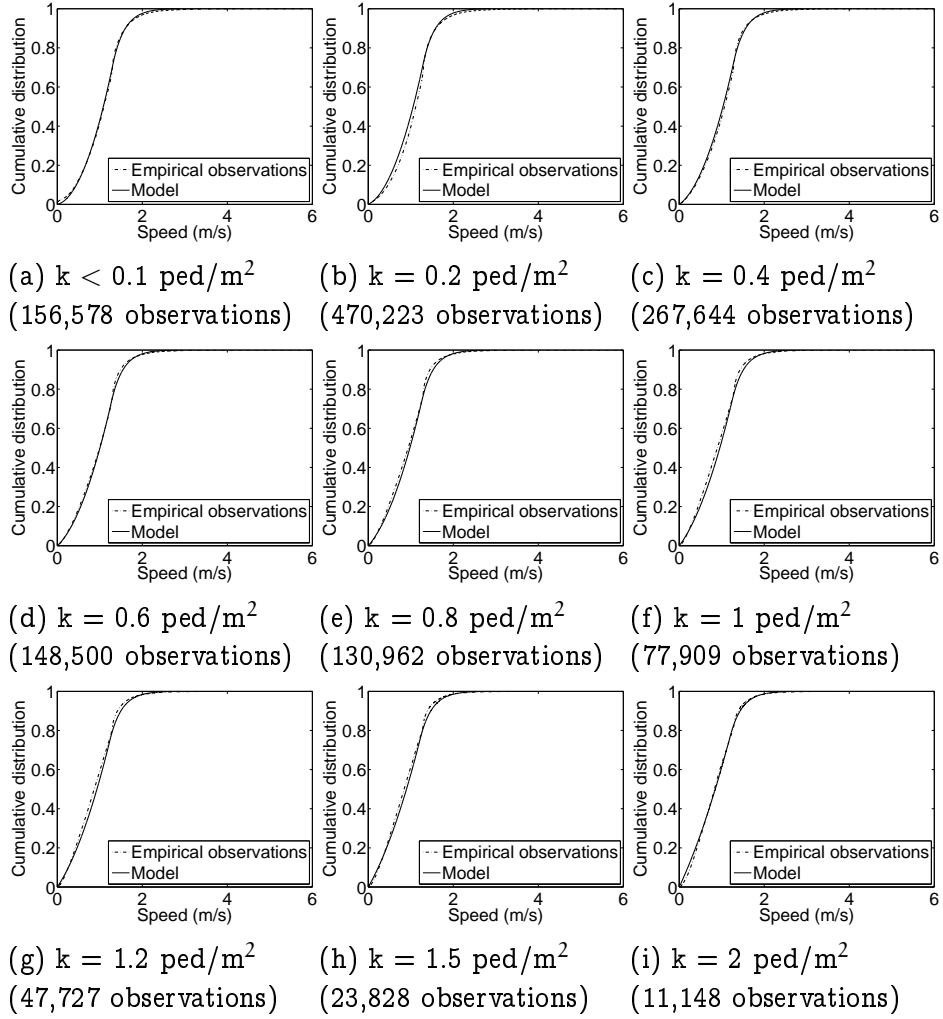
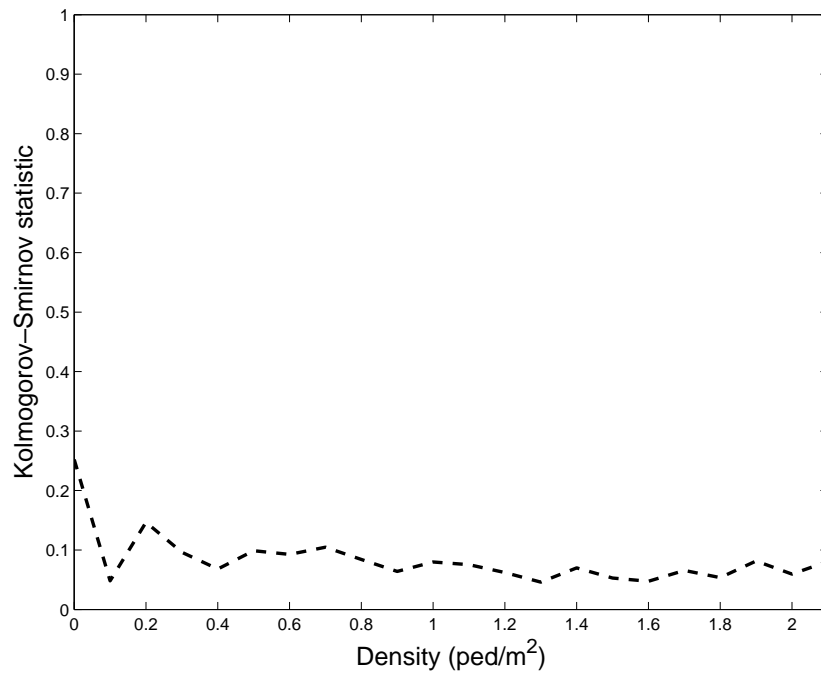
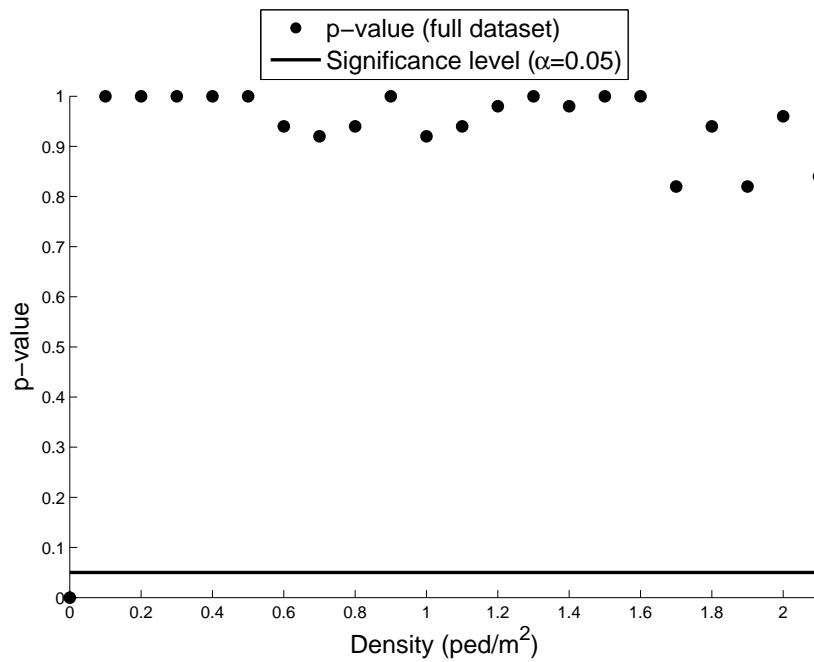


Figure 9: Comparison between model predictions (cumulative density) and empirical observations - Lausanne case study

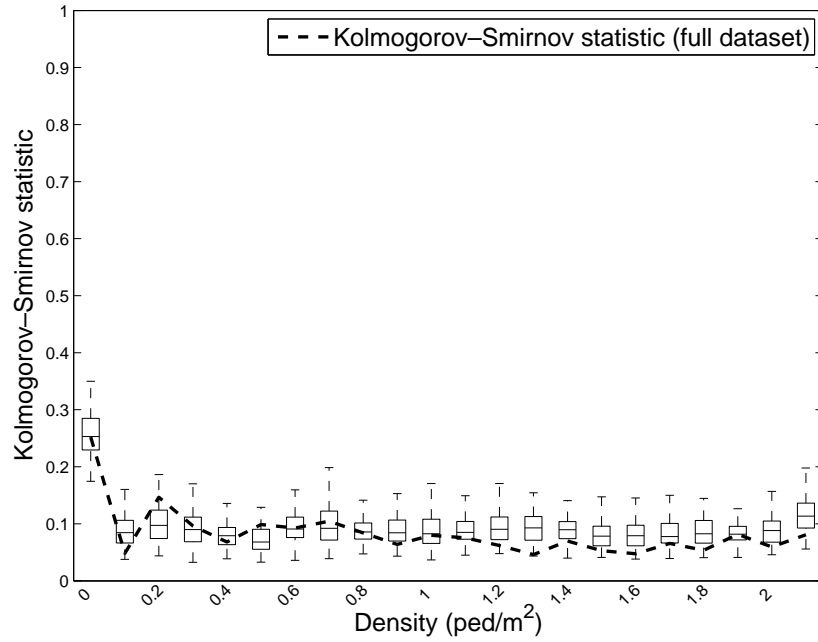


(a) Kolmogorov-Smirnov distance

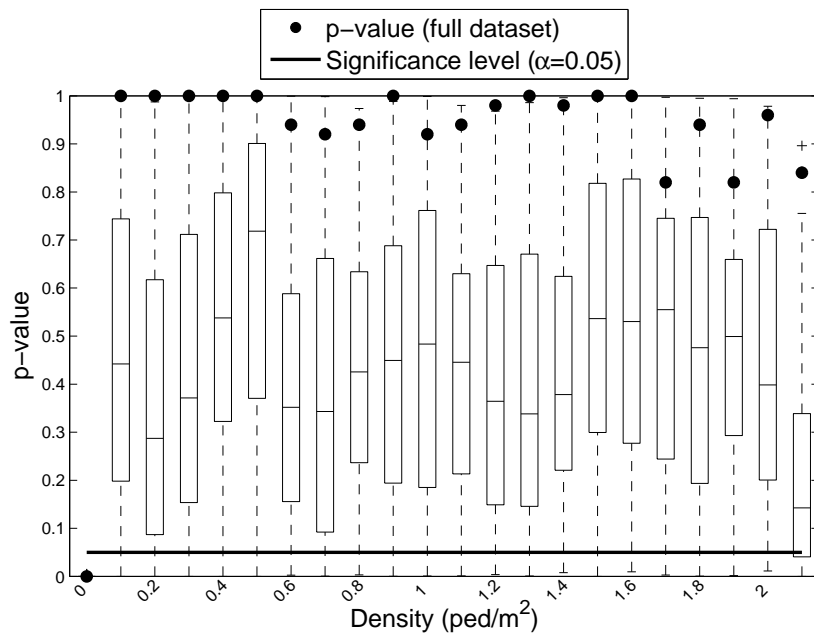


(b) p-value of Kolmogorov-Smirnov statistic

Figure 10: Kolmogorov-Smirnov validation - Lausanne case study



(a) Kolmogorov-Smirnov statistic



(b) p-value of Kolmogorov-Smirnov statistic

Figure 11: Specification test - Lausanne case study

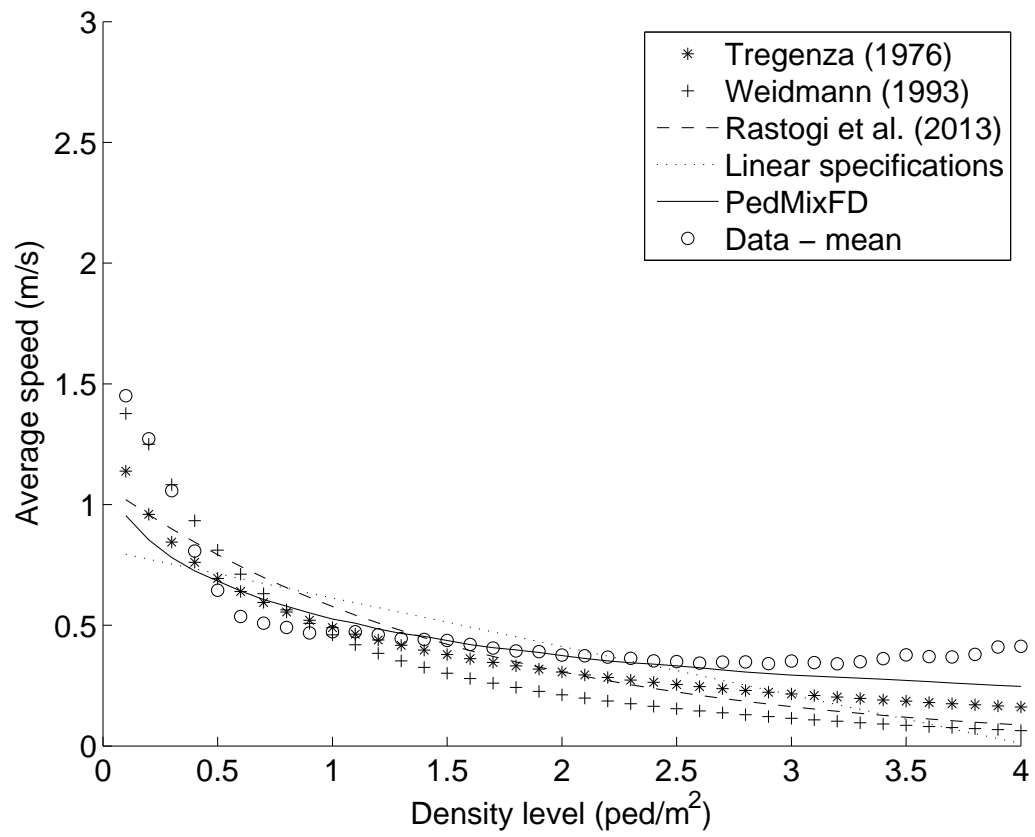


Figure 12: Comparison between deterministic models predictions and the aggregated probabilistic model predictions - Delft case study

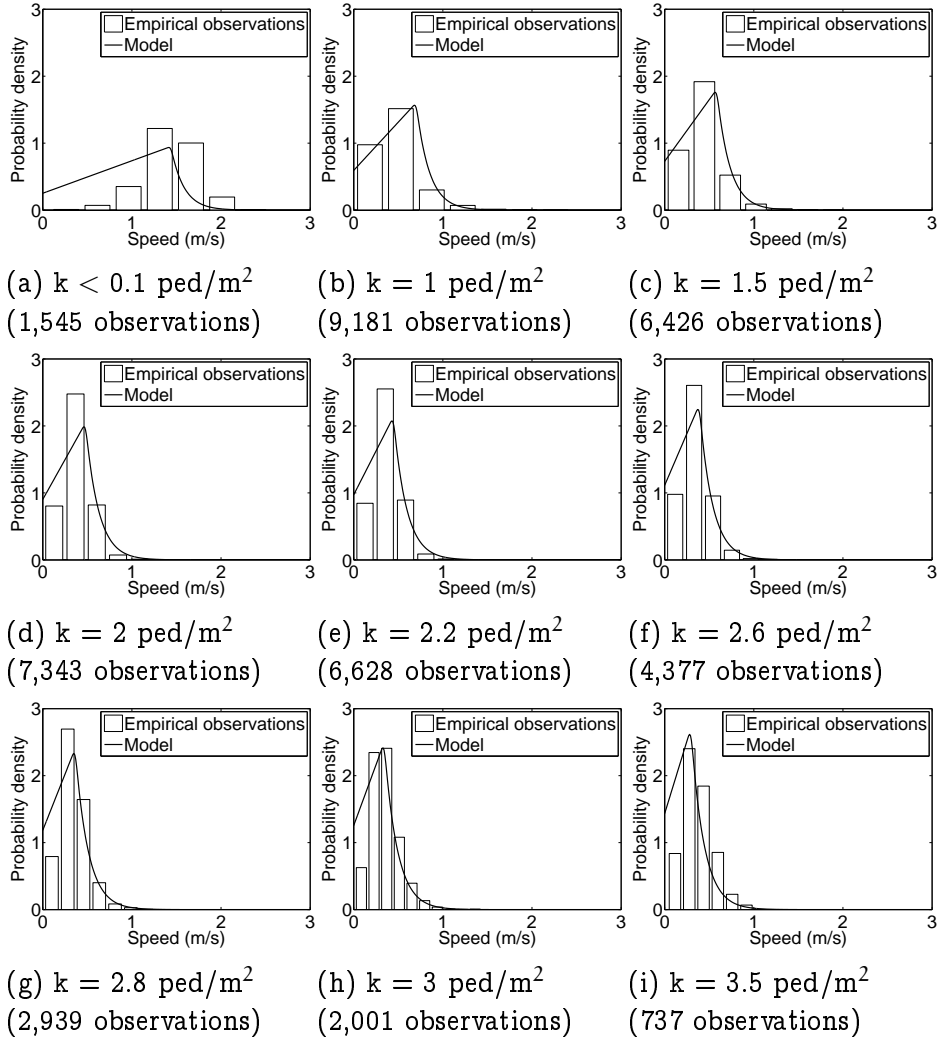


Figure 13: Comparison between model predictions (probability density) and empirical observations - Delft case study

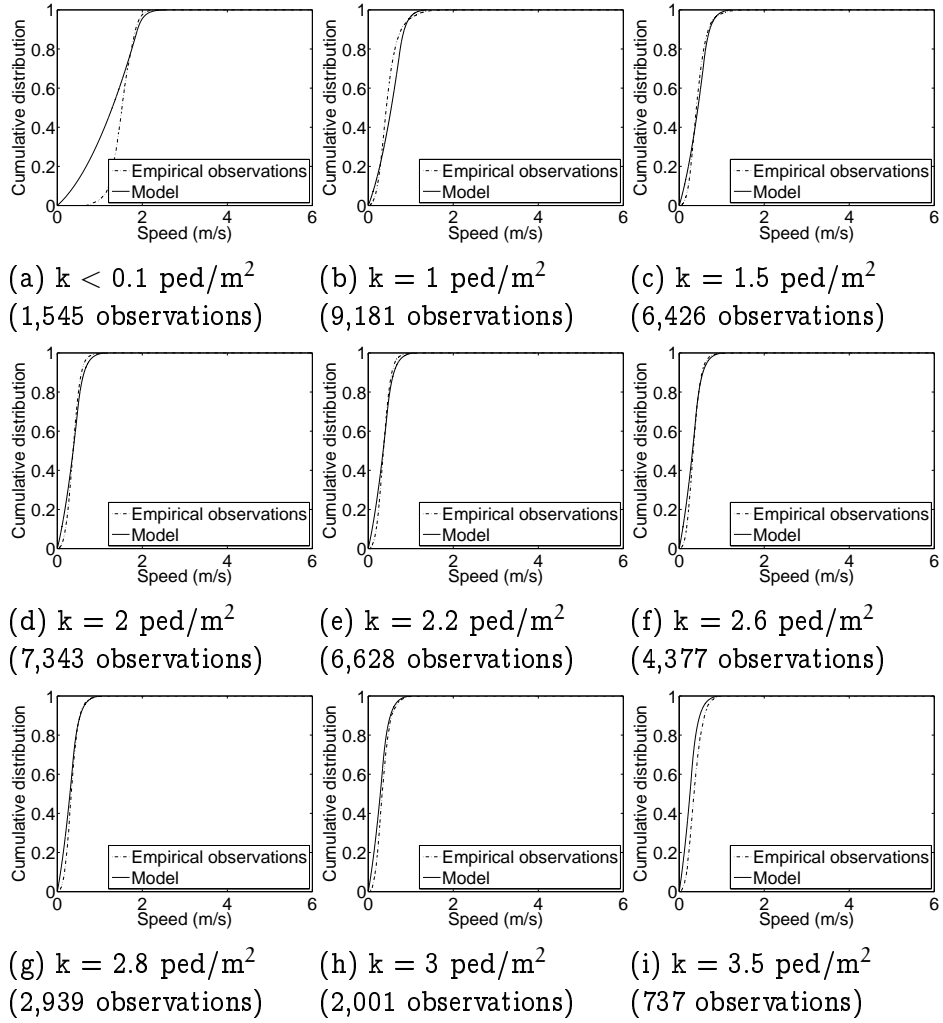
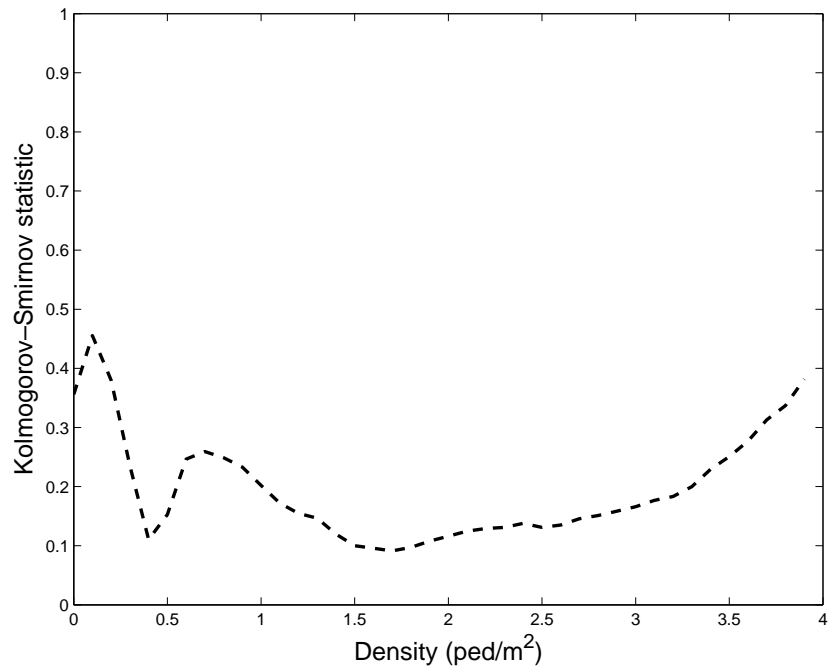
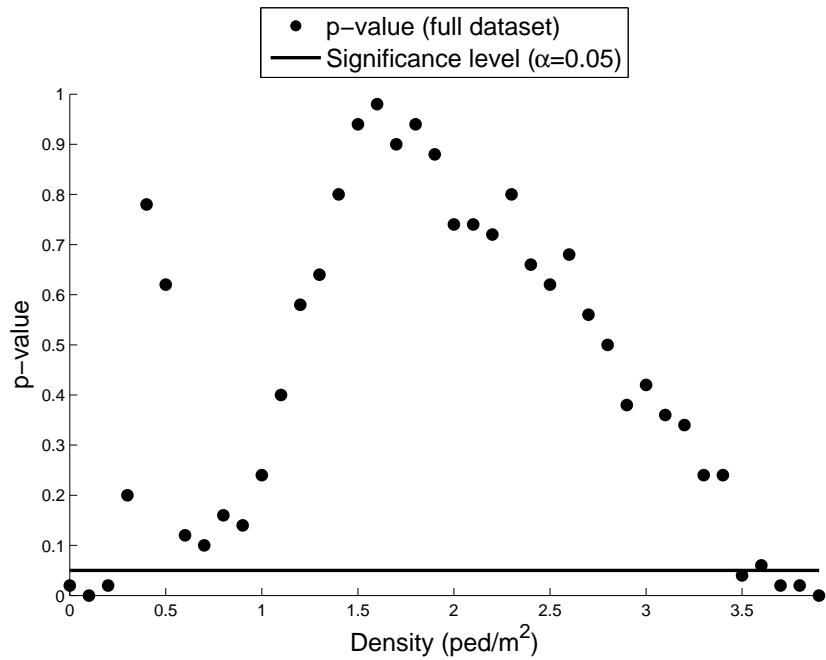


Figure 14: Comparison between model predictions (cumulative density) and empirical observations - Delft case study

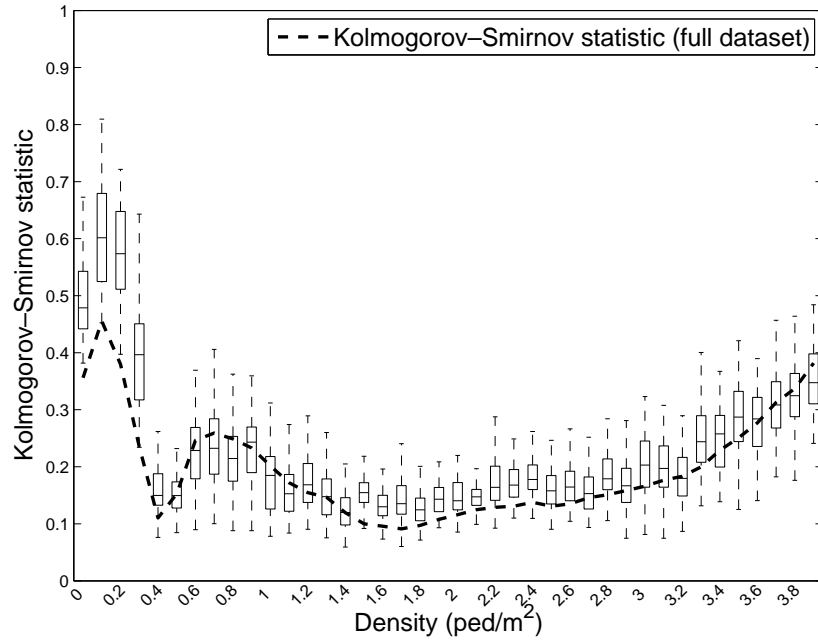


(a) Kolmogorov-Smirnov distance

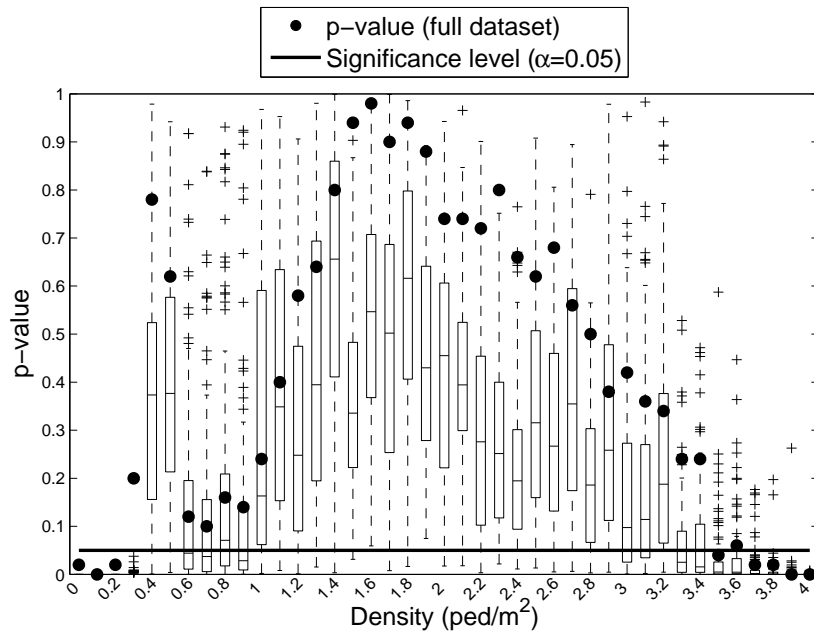


(b) p-value of Kolmogorov-Smirnov statistic

Figure 15: Kolmogorov-Smirnov validation - Delft case study



(a) Kolmogorov-Smirnov statistic



(b) p-value of Kolmogorov-Smirnov statistic

Figure 16: Specification test - Delft case study